

جامعة مؤتة

كلية الدراسات العليا

أثر نسبة القيم المفقودة وطريقة معالجتها في دقة تقدير معالم معادلة الانحدار البسيط

إعداد الطالب

عمران اسماعيل اللصاصمة

إشراف

الدكتور راجي عوض الصرايرة

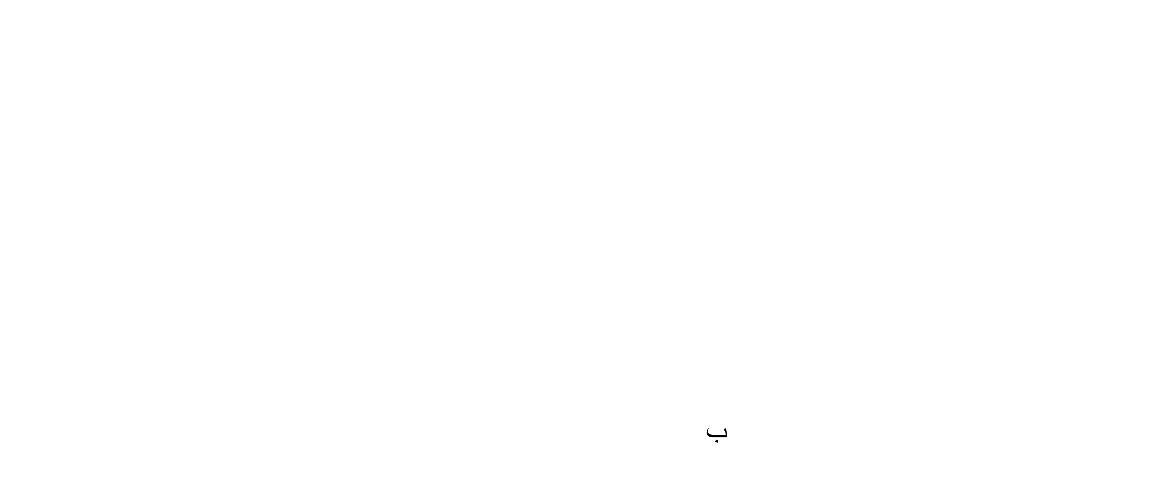
رسالة مقدمة إلى كلية الدراسات العليا

استكمالاً لمتطلبات الحصول على درجة الماجستير في

القياس والتقويم/ قسم علم النفس

جامعة مؤتة 2016

الآراء الواردة في الرسالة الجامعية لا تعبر بالضرورة عن وجهة نظر جامعة مؤتة



الإهداء

إلى من غرسوا في قلبي العلم والإيمان .

وعلموني معاني الحب والعطاء ... والوفاء والانتماء ... أمي وأبي.

إلى من اكتسبتُ منهم القوة والمحبة وعرفت معهم معنى الحياة.

إلى من تطلعوا لنجاحي بنظرات الأمل ... إلى قناديل دربي ... أخواني وأخواتي.

إلى من أسال أن يجعلهم ا قرة عين لي في الدنيا والآخرة وسترا من النار ... أبنائي وبناتي.

إلى من أكن لهم كل احترام وتقدير ... أصدقائي.

عمران اسماعيل اللصاصمة

الشكر والتقدير

الحمد رب العالمين والصلاة والسلام على خير البشرية محمد عليه افـضل الصلوات .

القائل:

" من لا يشكر الناس لا يشكر ا "

لا يسعني الا الوقوف والتقدم بوافر الشكر والتقدير إلى استاذي الفاضل الدكتور راجي الصرايرة على ملاحظاته القيمة وعلى الجهود العلمية الكبيرة التي بذلها فـي مراحل اعداد الرسالة وكان لجهوده هذه الأثر البالغ في اخراج الرسالة بالشكل التي هي عليه الأن .

كما اسجل خالص شكري وتقديري لأساتذة علـم الـنفس التربـوي / القيـاس والتقويم وزملائي طلبة الماجستير في تخصص القياس والتقويم.

عمران اسماعيل اللصاصمة

فهرس المحتويات

قائمة الجداول

قائمة الأشكال

الملخص

أثر نسبة القيم المفقودة وطريقة معالجتها في دقة تقدير معالم معادلة الانحدار البسيط

عمران إسماعيل اللصاصمة

جامعة مؤته، 2016

هدفت هذه الدراسة إلى معرفة أثر نسبة الفقد في القيم وطرق معالجتها على دقة تقدير معالم معادلة الانحدار البسيط، والمقارنة بين طرق المعالجة المستخدمة في الدراسة وتحديد افضل طرق المعالجة لهذه النسب المفقودة من البيانات، ولتحقيق هذا الهدف قام الباحث باستخدام برنامج(WINGEN3) لتوليد استجابات أختبارين مكون كل اختبار من (12) فقرة بمعامل صعوبة تباينه ثابت، موزعة على(200) مستجيب توزيعا طبيعيا، بوسط حسابي(0)، وانحراف معياري(1).

وقد اظهرت نتائج الدراسة أن نسبة الفقد (5%) لها تأثير على دقة معالم معادلة الانحدار البسيط عند اجراء مقارنة بين معادلة الانحدار البسيط الاصلية بدون فقد مع معادلة الانحدار البسيط بعد الفقد، ومقارنة مربعات البواقي لكل منهم، وكذلك نجد ان طريقة المعالجة بالتعويض المتعدد هي الافضل، لعدم وجود فروق تعود الى هذه الطريقة عند نسب الفقد المختلفة مما يؤدي الى عدم وجود تأثير في تقدير دقة معالم معادلة الانحدار البسيط. كذلك توصلت الدراسة الى أن طريقة المعالجة بالحذف لا ينصح باستخدامها لوجود تأثير في تقدير معالم معادلة الانحدار، كما توصلت الدراسة الى ان طريقة المعالجة بالوسط عند النسب (5%،10%) لا يوجد لها تأثير في تقدير دقة معالم معادلة الانحدار على العكس عند نسب الفقد(15%،20%) فانه يوجد تأثير على دقة التقدير لمعالم معادلة الانحدار.

Abstract

The impact of missing values ratio and the method of handling on accuracy of estimating simple regression equation parameters

Omran Ismail Allasasmeh

Mu'tah University, 2016

The aim of this study is to know the impact of the percentage of loss in values and processing methods on the accuracy of the estimation of parameters of regression equation is simple, and the comparison of methods used in the study and determine the best treatment methods for these percentages of lost data, in order to achieve this goal the researcher using the program (WINGEN3) to generate two responses each test component (12) paragraph difficulty coefficient variability, spread over (200) responsive natural distribution, Central algorithmic (0), standard deviation (1).

The results of the study showed that the percentage of loss (5%) have an effect on the accuracy of the parameters of the regression equation is simple when you compare the original simple regression equation without loss with simple regression equation after loss, compare residuals boxes each, as well as find that processing method with multiple recovery are better, the absence of differences back to the way when different loss ratios resulting in no impact to estimate the accuracy of the parameters of the regression equation. Also, the study found that treatment is not recommended for deletion used to having influence in the estimation of parameters of regression equation, the study also found that treatment center when the proportions (5%, 10%) have no impact on the estimation of the accuracy of the parameters of the regression equation on the contrary when the loss ratios (15%, 20%), there is an effect on the accuracy of the estimation of parameters of regression equation.

الفصل الأول

خلفية الدراسة وأهميتها

1.1 مقدمة:

إن التقدم والتطور العلمي المستمر والذي يعد من أهم روافد الحياة الأساسية لبناء وتقدم الأمم، لذا يسعى الباحثون في تجاربهم وأبحاثهم إلى استخدام مختلف النظريات والطرق الإحصائية والقوانين للحصول على نتائج دقيقة في أبحاثهم من أجل إيجاد الحلول المناسبة لكل العقبات التي قد تعرقل مسيرتهم البحثية إلا أنهم لاحظوا وجود عقبة وهي القيم المفقودة في البيانات المراد تحليلها ومعالجتها.

حيث أشار الكثير من الباحثين أن هناك مشكلة تواجههم عند قيامهم بجمع وتحليل البيانات وهي عدم اكتمال الإجابات أو البيانات حيث تبدأ مع الباحث أثناء التطبيق وحتى جمع وتصحيح الاستجابات، على أداة جمع البيانات فالمتوقع هو استجابة كل أفراد العينة بدقة على كل فقرة من فقرات البحث وتسجيل استجاباتهم بدقة. هذه هي الحالة المثالية إلا أن هذه الحالة لا تظهر في كثير من الأبحاث، فالبيانات التي يتم جمعها في اغلب الاحيان تكون ناقصة وتؤدي إلى تقديرات متحيزة واقل كفاءة (Little & Rubin,1987). فالغاية من الاختبارات هي الحصول على معلومات مهمة ولكنها قد تكون غير دقيقة نتيجة وجود بيانات مفقودة.

وقد اشار ليودلو و أوليري (Ludlow & OLEARY,1999) أن سبب وجود القيم المفقودة هو عدم الوصول إلى بعض الفقرات لعدم وجود الوقت الكافي أو لعدم الاهتمام من المستجيب، أيضا أن عدم إجابة المستجيب لبعض الفقرات بغير قصد أو لعدم قدرته على الإجابة يؤدي إلى وجود القيم المفقودة.

وقد ذكر ميكانيت وميكانيت وسيداني وفيغورد (McKnight, Sidani & Mcknight, Figueredo, 2007) ثلاثة فئات لأسباب الفقد للبيانات وهي أولاً أسباب تعود إلى فئة المستجيبين أنفسهم حول خاصية معينة تتعلق بهم مثلاً الدخل الشهري وثانياً منها أسباب تعود لتصميم الدراسة؛ أي أنها تحتاج لوقت طويل من المستجيب للإجابة عليها وأخيراً أسباب تعود للتفاعل بين المستجيبين وتصميم الدراسة.

ويشير سكفر واجراهام (Schafer & Graham, 2002) أن مطورو البرامج

الحاسوبية يبذلون جهود كبيرة لمعالجة مشكلة البيانات المفقودة وإعطائهـا مظهـر الاكتمال. ولكن يمكن أن يسبب تحرير البيانات ضررا أكبر من نفعه بإنتاجه أجوبة منحازة وغير فعالة، وغير موثوقة.

وقد ذكر بيو وأندرس (Peugh & Enders 2004) إلى أن كثير من الباحثين يعالجون القيم المفقودة في أبحاثهم بالإهمال. على الرغم من أن هذه القيم قد تكـون لها أهمية بتغير نتائج البحث. الأمر الذي يؤدي إلى نتائج غير دقيقـة، لـذا يجـب معالجة القيم المفقودة بطرائق التعويض المناسبة.

ويجدر الاشارة إلى أن العلماء صنفوا طرق التعامل مع البيانات المفقودة ومعالجتها إلى طرق قائمة على الحذف (Methods Depend on Deletion) وطرق قائمة على احتساب قيمة تعويضية (Imputation) للقيم المفقودة.

ونظرا لأهمية القرارات التي تبنى على نتائج معادلة الانحدار الذي يدرس العلاقة بين المتغيرات برزت أهمية الدراسة لبحث أثر نسبة البيانات المفقودة وطرق معالجتها على دقة معالم معادلة الانحدار البسيط من أجل تحديد أفضل طرق المعالجة التي ستستخدم في الدراسة.

2.1 مشكلة الدراسة:

لا تخلو بيانات الدراسات والبحوث التربوية من مشكلة وجود قيم مفقودة في بياناتها، فغالبا ما يتم التعامل مع هذه القيم بالتجاهل الأمر الذي بدوره قد يؤثر على دقة نتائج هذه الدراسات. ومن هذه الدراسات ما يعنى ببناء الاختبارات وتطويرها في ضوء نظرية استجابة الفقرة والنظرية الكلاسيكية التي تتطلب عينات بحجوم كبيرة لما لها من دور في زيادة الدقة في تقدير معالم معادلة الانحدار ووجود قيم مفقودة في هذه البيانات قد يؤثر على دقة نتائجها. من هنا برزت العديد من الطرق للتعامل مع القيم المفقودة التي تظهر البيانات كاملة دون نقص مما قد يعمل على زيادة موثوقية النتائج بتلك الدراسات.

كما أن تأثير هذه القيم يعتمد على حجمها قياسا مع حجم البيانات الاصلية ،فقد أوضح جراهام(Graham ,2009) أن زيادة نسبة الفقد في البيانات عن (5%)

فإن تأثيرها على نتائج هذه الدراسات يزداد. فمن هنا جاءت الحاجة إلى هذه الدراسة والتي ستتناول أثر نسبة الفقد في البيانات على دقة معالم معادلة الانحدار البسيط.

3.1 أسئلة الدراسة:

تتمحور أسئلة الدراسة بالآتي:

السؤال الأول: ما أثر طرق معالجة البيانات المفقودة (الحذف، حساب المتوسط، التعويض المتعدد). على دقة معاملات معادلة الانحدار البسيط ؟

السؤال الثاني: ما أثر نسب الفقد للبيانات (5%،10%،15%،20%). على دقة معاملات معادلة الانحدار البسيط؟

السؤال الثالث: هل توجد فروق بين المتوسطات الحسابية لمربعات البواقي لتقدير درجات الأفراد تعزى لنسب الفقد(5%،10%،15%،20%).

السؤال الرابع: هل توجد فروق بين المتوسطات الحسابية لمربعات البواقي لتقدير درجات الأفراد تعزى لطرق معالجة البيانات المفقودة (الحذف، حساب المتوسط الحسابي، التعويض المتعدد).

4.1 أهداف الدراسة:

هدفت هذه الدراسة إلى الكشف عن أثر طرق معالجة البيانات المفقودة على دقة معاملات معادلة الانحدار البسيط، واثر نسب الفقد للبيانات على دقة معاملات معادلة الانحدار البسيط، وهدفت ايضا إلى الكشف عن وجود فروق بين المتوسطات الحسابية لمربعات البواقي تعزى لنسب الفقد وطرق معالجتها.

5.1 أهمية الدراســة:

تنبع أهمية الدراسة في جانبين أولهما الجانب النظري والذي يتمثل في توضيح مفهوم القيم المفقودة وطرق معالجتها وكيفية الاستفادة من اسلوب المحاكاة في معالجة مشكلة القيم المفقودة، وثانيهما هو الجانب العملي وتمثل ذلك في اعطاء نتائج تتعلق في اثر نسبة القيم المفقودة وطرق معالجتها على دقة تقدير معالم معادلة

الانحدار والتي تعد دليل عملي لكل الباحثين عندما يتم استخدام تحليل الانحدار البسيط في كيفية تعاملهم مع مشكلة القيم المفقودة.

6.1 حدود الدراسـة:

1. تحديد نسبة الفقد في البيانات ب (5%، 10%، 15%، 20%) من مجموع البيانات المولدة.
2. اقتصرت الدراسة على تناول ثلاث طرق في معالجة القيم المفقودة وهي (الحذف، التعويض بالوسط الحسابي، وحساب القيم التعويضية المتعددة).
3. اقتصرت هذه الدراسة على البيانات التي تم توليدها باستخدام برمجية (Wingen3).

7.1 مفاهيم الدراسة وتعريفاتها الاجرائية:

1. القيم المفقودة (Missing Values): عدم استجابة المستجيب عـن بعـض فقرات الاختبار أو الاستبيان وترك هذه الفقرات فارغة دون إجابة بـصرف النظر عن سـبب ذلـك، ويعبـر عـن نـسبة الفقـد مـن خـلال فقـد (5%،10%،15%،20%) من القيم.
2. (القيمة التعويضية): هي القيمة التي يتم وضعها مكان القيمة المفقودة بعد أن يتم معالجتها بواحدة من طرق معالجة القيم المفقـودة(الحـذف ، حـساب الوسط، التعويض المتعدد)
3. دقة التقدير (Accuracy of Estimate):تشير إلى جودة التقـدير للمعـالم والقدرات، التي تتميز بالاحتمال الكبير في أن يكون التقدير اقرب من القيمة الحقيقية للمعلم أو القدرة، وذلك باختيار التقدير غير المتحيـز Unbiased Estimator والذي يكون صاحب اقل تباين من التقديرات الاخـرى غيـر المتحيزة، وذلك باستخدام الخطأ المعياري في التقدير، أو معـدل مربعـات الاخطاء أو البواقي.
4. تقدير المعالم (Parameters Estimate): تعبير كمي عن المعـالم التـي

تعتمد عليها التوزيعات الإحصائية للاستجابة علـى فقـرات الاسـتبيان أو الاختبار بمقاييس مستخلصه من البيانات بحيـث تتـضمن هـذه المقـاييس تقديرات تمثل خصائص المجتمع (Population)

5. مقارنة طرق معالجة البيانات: اسلوب يمكن من خلاله معرفة الطريقة الأكثر فاعلية في معالجتها للقيم المفقودة، وذلك من خلال استخدام الخطأ المعياري للقيمة المقدرة حيث تتناسب فاعلية الطريقـة تناسـبا عكـسيا مـع الخطـأ المعياري، ويستحب أن يكون الفرق بين القيمتين المقدرة والحقيقية اقل مـا يمكن.
6. البيانات المولدة (Generated Data): البيانات المستخدمة في هذه الدراسة والمولدة بواسطة برنامج توليد البيانات (WINGEN3)
7. نموذج الانحدار البـسيط(Simple linear regression) : هـو نمـوذج إحصائي يقوم بتقدير العلاقة التي تربط متغير كمي واحد وهو المتغير التابع مع متغير كمي أخر وهو المتغير المستقل. وينتج من هذا النمـوذج معادلـة إحصائية خطية يمكن استخدامها لتفسير العلاقة بين المتغيرين أو تقدير قيمة المتغير التابع عند معرفة قيمة المتغير المستقل. ويمكـن صـياغة العلاقـة بالنموذج التالي :

$$\widehat{yi} = \hat{a} + bx_i + \hat{e}$$

حيث أن:

$\widehat{yi}$: هو المتغير التابع

x_i: هو المتغير المستقل

$\hat{e}$: هو الخطأ العشوائي

$\hat{a}$: هي قيمة ثابته تعبر عن قيمة $\widehat{yi}$ عندما تكون قيمة x_i تساوي صفر وقيمة b تعبر عن ميل الخط المستقيم الذي يوضح العلاقة.

الفصل الثاني

الإطار النظري والدراسات السابقة

1.2 الإطار النظري:

عند جمع البيانات قد يكون هناك بيانات مفقودة أو عدم وضوح لهذه البيانات، وكون هناك مقاييس تعتمد على عدد هذه القيم فإن ذلك يؤدي إلى مشكلة بسبب وجود هذه البيانات المفقودة، فلذا يتوجب علينا التفكير في إيجاد الحلول المناسبة لمثل هذه الحالات(هاوت دينس، كرايمر دانكان، 2016).

ويضيف مسلفي و ويو (Mislevy & Wu,1988) إن إجابات الأفـراد فـي أي وقت على اختبـتار أو استبيان تكون مفقـودة لسببيـن، إما باختيـار المـستجيب نفسه أو بسبب تصميم الاختبار، ويمكن التفريـق بـين نـوعين لعـدم الاسـتجابة (Huisman,Korl & Vansonderem, 1998)

الاول: يتعلق بعدم الاستجابة الكاملة لجميع الفقرات: لا يستجيب الفرد في هذه الحالة لأي فقرة من فقرات الاختبار أو رفضه للمشاركة لمخاوف حول سرية المعلومات أو عدم القدرة على التواجد بسبب العمر أو الحالة الصحية.

والنوع الثاني: يرتبط بعدم الاستجابة للفقرة: أن المستجيب يشارك في الاختبار فيجيب على بعض الفقرات ويترك بعضها دون الاجابة، مما يوجد بيانات جزئية غير كاملة بسبب فقدان بعضها وتشمل: تخطي المستجيب للفقرة اي تركها دون اجابه، وذلك لأنه تخطاها بدون قصد أو لأن الزمن غير كافي للإجابة، أو لأنه لا يعرف الاجابة لسبب ما، أو لأنها تحتاج إلى مزيد من الجهد، أو لأن بعض المستجيبين لا يجيبوا على بعض الفقرات لأنها تتعلق بأمور خاصة بهم. والبعض الاخر لا يجيب على بعض الفقرات لأنها تعالج موضوعا حساساً، أو لأن الفقرة المعروضة غير مناسبة للمستجيبين من حيث المضمون أو لا تنطبق عليهم. أو لأن بعض المستجيبين لا يستخدم التخمين للفقرات عندما لا يستطيع الاجابة عليها، أو لطول الاختبار الذي يرهق المستجيبين، ويؤثر على تركيزهم ومستوى تدفق المعلومات لديهم، فظروف

المقياس تفرض نفسها أحيانا، وتؤثر على افراد العينة فالبعض ينسحب والاخر يعتذر.

وقد ذكر (الزعبي، 2013) إلى انه توجد عدة طرق للتعامل مع القيم المفقودة ،حيث يساعد معرفة الباحث للنمط (Pattern) التي تظهر عليه القيم المفقودة، وكذلك معرفته لآلية الفقد (Mechanism) بسبب الفقد –على اختيار الطريقة المناسبة للتعامل مع القيم المفقودة. و(المقصود بالنمط) هو كيفية ظهور القيم المفقودة في مجموعة البيانات- موقع ظهور القيم المفقودة- وقد ميز ليتل وروبين Little & Rubin ,2002) بين ثلاثة أنواع من أنماط فقد القيم هي: النمط الافتراضي أو الاعتباطي(Arbitrary Pattern)، والنمط وحيد المتغير (Univariate Pattern)، والنمط الوتيري (Monotone Pattern)، في حين أشار أندرس (Enders ,2010) إلى أن الادب السابق ميز بين ستة أنواع من أنماط فقد القيم هي:

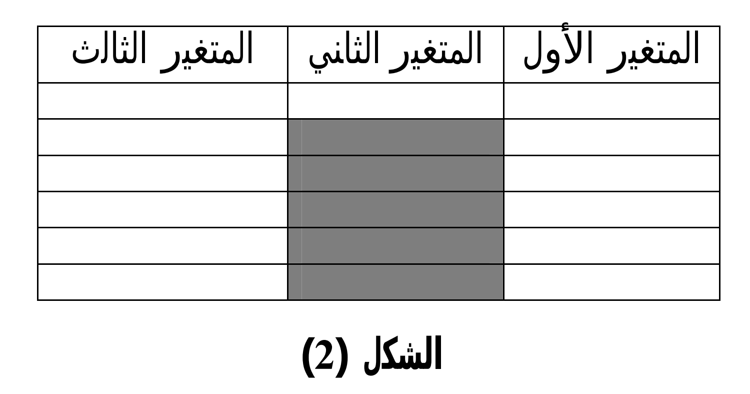
1. النمط الاعتباطي (Arbitrary Pattern) النمط العام (General Pattern) في هذا النمط تكون القيم المفقودة منتشرة بشكل عشوائي (دون شكل معين).

المتغير الأول	المتغير الثاني	المتغير الثالث
	■	
■		■
	■	
		■
		■

الشكل(1)

النمط الافتراضي. المنطقة المضللة تشير إلى أماكن البيانات المفقودة

ب. النمط وحيد المتغير (Univariate Pattern): ويحدث هذا النمط مـثلا عنـدما تكون هناك فقرة في الاختبار أو الاستبانة لها حساسية عند بعض الأفراد. بمعنى أن القيم المفقودة متعلقة بفقرة واحدة فقط من فقرات المقياس أو الاختبار(متغير واحد)، إذ يوجد عدد من المستجيبين لم يستجيبوا على تلك الفقرة بينمـا بـاقي الفقرات أو المتغيرات تحوي بيانات كاملة.

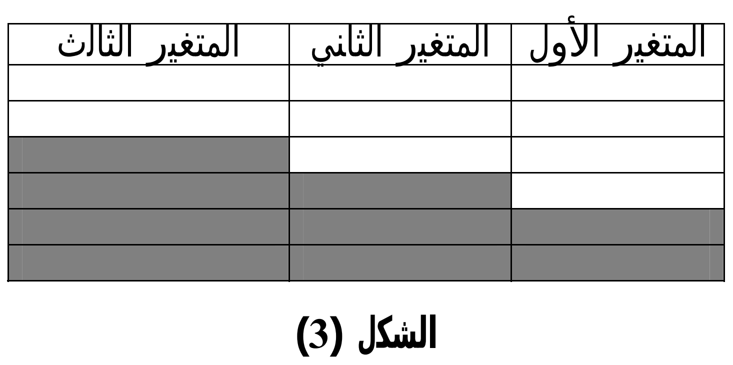


المتغير الأول	المتغير الثاني	المتغير الثالث

الشكل (2)

النمط وحيد المتغير والمنطقة المضللة تشير إلى أماكن البيانات المفقودة

ج. النمط الوتيري (Monotone Pattern): حيث يظهر في هذا النمط أثر الهدر لأفراد العينة إذ يقرر بعض الأفراد الانسحاب من الدراسة بعد المرحلة الأولى ثم يقرر أفراد أخرون الانسحاب بعد المرحلة الثانية وهكذا، فتظهر القيم المفقودة على شكل درج بحيث أن القيم المفقودة تزداد مع ازدياد المرحلة أو مع صعوبة الفقرة، بحيث إذا كانت الفقرة (i) مفقودة فإن الفقرة (I+1 , I +2 , … , n) تكون مفقودة أيضاً.

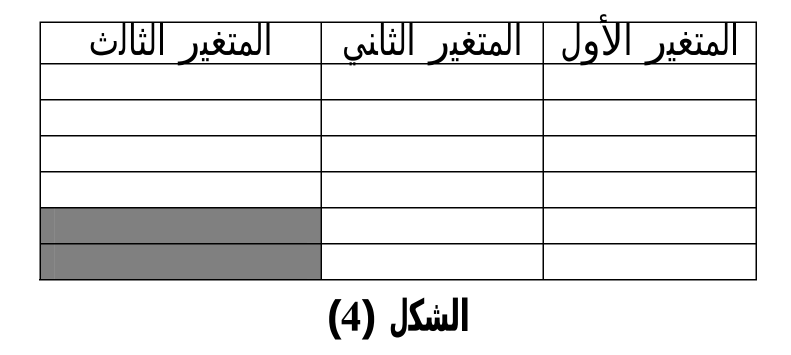


المتغير الأول	المتغير الثاني	المتغير الثالث

الشكل (3)

النمط الوتيري والمنطقة المضللة تشير الى أماكن البيانات المفقودة

د. نمط وحدة عدم الاستجابة (Unit Nonresponse Pattern): بمعنى أنه لو كان هناك ثلاثة متغيرات إثنان منها متوفر بياناتها لجميع المستجيبين، والمتغير الاخر يرفض بعض المستجيبين الاجابة عليه، وغالبا ما يحدث هذا النمط في البحوث المسحية.



المتغير الأول	المتغير الثاني	المتغير الثالث

الشكل (4)

نمط وحدة عدم الاستجابة والمنطقة المضللة تشير إلى أماكن البيانات المفقودة

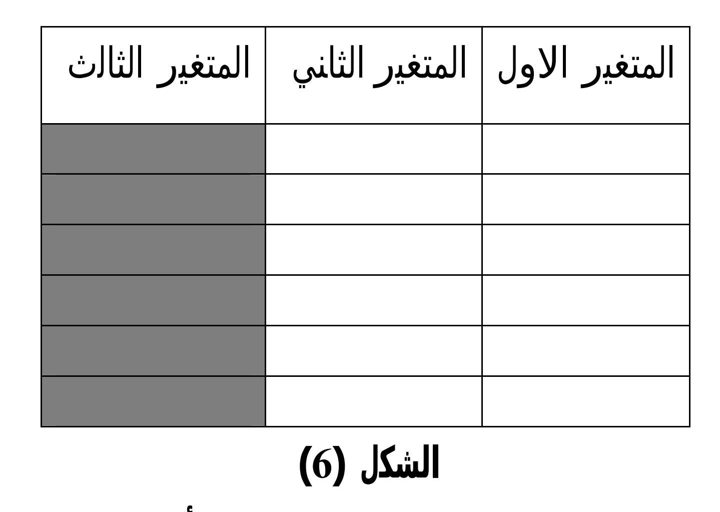
ه. نمط البيانات المفقودة المخطط لها (The planned Missing Data Pattern): حيث يتم التخطيط لهذا الفقد من قبل الباحثين عند عملية جمع البيانات وتجهيز ادوات الدراسة ويعتبر هذا النمط ذو فائدة عند جمع بيانات مقياس يتضمن عدد كبير من الفقرات حيث يتم تقسيم فقرات المقياس إلى ثلاثة اجزاء مثلا، وتشكيل ثلاث صور للمقياس كل صورة تحتوي على جزء معين .

المتغير الأول	المتغير الثاني	المتغير الثالث

الشكل (5)

نمط البيانات المفقودة المخطط لها والمنطقة المضللة تشير إلى أماكن البيانات المفقودة

و) نمط المتغير الكامن(Latent Variable Pattern): في هذا الـنمط تفقـد البيانـات بسبب متغير كامن ولجميع المستجيبين بالرغم أنه ليس مـن الـضروري عـرض المتغير الكامن في نماذج التحليل المستخدمة كمشكلة فقد البيانات حيث تبنى الباحثون خوارزميات (Algorithms) لفقد البيانات لتقدير هذه النماذج علـى سـبيل المثـال: نمـوذج متعـدد المـستويات (Multilevel Models; Raudenbush & Bryk,2002,PP.440-444) .



المتغير الاول	المتغير الثاني	المتغير الثالث

الشكل (6)

نمط المتغير الكامن والمنطقة المضللة تشير إلى أماكن البيانات المفقودة

أما آلية فقد القيم (Missing Values Mechanisms) فتعني الكيفية التي تمت من خلالها عملية الفقد لبعض القيم من بعض متغيرات الدراسة، حيث يكون الاهتمام بأسباب عدم الاستجابة والعلاقة بين هذه الأسباب والمتغيرات المساعدة. وتحدد آليات فقد القيم العلاقات الاحتمالية بين الفقد والقيم المشاهدة في مجموعة البيانات. وقد قسم ليتل وروبين (Little & rubin,2002) أليات فقد القيم إلى ثلاثة اقسام هي:

أ- الفقد العـشوائي بالكامـل (Missing Completely At Random)(MCAR): وتكون القيم المفقودة عشوائيا بالكامل عندما تكـون عمليـة الفقـد بـسبب العشوائية المحصنة، بمعنى أن الفقد لا يرتبط بأي متغيـر مـن متغيـرات الدراسة وغير مرتبط بقيم المتغير نفسه، أو بالظروف التجريبيـة للدراسـة، وإنما بسبب الصدفة التامة .فمثلاً يمكن صدفة أن يترك مستجيب فقرة ما في اختبار موضوعي ويمكن التأكد من الفقد العشوائي الكامل للبيانات من خلال مجموعة من الاختبارات، مثل اختبار ليتل (Little,1988) المتوفر في برنامج الرزمة الإحصائية للعلوم الاجتماعية (SPSS).

ب- الفقد العشوائي (Missing At Random)(MAR): وفي هـذا النـوع تتـأثر البيانات المفقودة بخصائص الأفراد المستجيبون ولا تتأثر بخصائص البيانات المفقودة نفسها، ومثال ذلك إذا كان مضمون الفقرة يتعلـق بمقـدار الـدخل الشهري للفرد المستجيب، وكانت هذه الفقرة مفقودة (غير مجابه) مـن قبـل

بعض الأفراد المستجيبين، فإن السبب في الفقد في هذه الحالة لا يعزى للقـيم الحقيقية للدخل، ولكن ربما يعزى لمتغير اخر مثل المستوى التعليمي لهؤلاء المستجيبين الذين لم يقوموا بالإجابة عن تلك الفقرات، وكذلك الأفراد أصحاب المؤهلات العلمية العالية، ربما يميلون لعدم الافصاح أو الكشف عن مقـدار دخولهم، بالمقارنة مع الأفراد أصحاب المؤهلات العلمية المتدنية.

ج- الفقد غير العشوائي (Missing Not At Random)(MNAR):وفي هذا النـوع تكون هناك علاقة بين القيم المفقودة ومتغيـر الدراسـة، بمعنـى احتمـال الاستجابة يعتمد على القيم المفقودة، ولا يمكن تفسيره من خلال البيانات التي تم قياسها لجميع عناصر العينة، ولا يوجد اختبار إحصائي لتحديد إذا كانـت القيم مفقودة عشوائياً (MAR) أو غير عشوائياً (MNAR)، إذ أن اختبار ليتل فقط يحدد فيما إذا كانت القيم مفقودة عشوائيا بالكامل .

وبغض النظر عن نوع آلية الفقد في البيانات، فإن القيم المفقودة تخلق مـشاكل في نتائج الدراسة وتفسيراتها، الأمر الذي تنبه له الباحثون وبحثوا في عملية التعامل معها ومعالجتها بالشكل الأمثل.

وقد شاع خلال العقدين الاخرين موضوع التعامل مع البيانات المفقودة، بحيـث طرا تقدم ملموس وتحسن في الأساليب والإجراءات الإحصائية التي تعالج البيانـات المفقودة وما ينتج عنها من مشاكل وتحيزات، ومع ذلك فمـا يـزال سـوء الفهـم والممارسة غير الـسليمة سـائدين لـذا يلجـأ البـاحثون إلـى الطـرق القديمـة للتعامل مع البيانات المفقودة والتي تسبب في معظم الأحيان إلـى التحيـزات فـي النتائج (Schafer & Graham,2002).

لهذا يبذل مطورو البرامج الحاسوبية جهود كبيرة لمعالجـة مـشكلة البيانـات المفقـودة وإعطائها مظهـر الاكتمال. ولكـن يمكن أن يـسبب تحريـر البيانـات ضرراً أكبـر من نفعـه بإنتاجـه أجوبة منحازة وغير فعالـ ـة، وغيـر موثوقـة (Schafer & Graham 2002). وقد أشار بيو وأندرس (Peugh & Enders 2004) إلى أن كثير من الباحثين يواجهون القيم المفقودة في أبحاثهم إما بالإهمال أو إعطائها قليل من الانتباه. على الرغم أن هذه القيم قد تكون لها أهمية ما يغير نتائج البحـث.

وفي أغلب الأحيان قد تفرض القيم المفقودة نفسها على التحليل الإحصائي مما يؤدي إلى نتائج غير دقيقة، لذا يجب معالجة القيم المفقودة بطرائق التعويض المناسبة.

ويمكن معالجة البيانات المفقودة وتصنيفها إلى: طرق قائمة على الحذف (Methods Depend on Deletion)، وطرق قائمة على احتساب قيمة تعويضية (Imputation) للقيمة المفقودة. ويندرج تحت كل طريقة عدة طرق للحذف منها (حذف الحالة، إعادة الوزن، الخطأ، غير الموجودة). أما الطريقة التعويضية فهي تعتمد على التقدير (Estimation) لتعويض القيم المفقودة وهي (حساب قيمة تعويضية واحدة، حساب قيمة تعويضية واحدة من خلال الانحدار، حساب قيمة تعويضية للوسط المصحح للفقرة، حساب قيمة تعويضية بطريقة دالة الاستجابة، خوارزمية تعظيم التوقعات، حساب قيم تعويضية متعددة، الطريقة الصحيحة جزئيا.

وقد تم في هذه الدراسة استخدام ثلاث طرق للمعالجة وهي:

1- حذف الحالة (Case Deletion): وتقوم هذه الطريقة على إنهاء عملية التعامل مع القيم المفقودة قبل البدء بعملية التحليل (Witta & Kaiser,1991)، وتنقسم هذه الطريقة من حيث التحليل إلى إسلوبين، هما حذف الحالة الكاملة (List wise Deletion)، وحذف الحالة المتوفرة (Available- case Analysis) وهما موجودان كخيارين من خيارات أخرى في برنامج (SPSS)، ففي طريقة الحذف الكاملة، يتم حذف أي مستجيب له قيمة مفقودة لأي فقرة من فقرات الاختبار أو المقياس، بمعنى تحليل بيانات المستجيبين الذين لديهم استجابات كاملة على جميع الفقرات. وهذا الحذف قد يؤثر على قضية تمثيل العينة للمجتمع خاصة إذا كان عدد المستجيبين الذين تم حذفهم كبير، حتى لو كان عدد المستجيبين الذين تم حذفهم قبل التحليل قليل، فإن ذلك قد يؤدي إلى خسارة للمعلومات بسبب حذف بعض المستجيبين وبالتالي يزداد احتمال التحيز ويمكن أن تكون هذه الطريقة غير فعالة في العينات الصغيرة. ولطريقة حذف الحالة الكاملة محددات رئيسية، تتمثل في أنها تتطلب إفتراض أن تكون البيانات مفقودة عشوائيا بالكامل(MCAR)، والا فإن هذه الطريقة تؤدي إلى نتائج متحيزة بسبب أن باقي العينة لا تمثل المجتمع الأصلي. كما أنها تقلل من حجم العينة المتوفرة، حيث أن تخفيض حجم العينة نتيجة الحذف يؤثر

سلبا في القوة الإحصائية للاختبار، وبالتالي اضعاف قدرة الباحث على تحديد الأثـر الدال من الاختبار الإحصائي(Gemici,Bednarz,&LIM,2012).

كما يذكر فيغاردو وميكنايت وسيداني (Figueredo, McKnight, McKnight & Sidani,2000) وجراهام (Graham,2009)، أن بعض الدراسات أشـارت إلـى أن طريقة الحذف الكاملة تعمل بشكل جيد عندما تكون آلية فقد البيانات تتبع آلية الفقـد العشوائي الكامل (MCAR)، وعندما يكون حجم العينة كبيرا، بحيث أن التخاص من القيم المفقودة لا يقلل من القوة الإحصائية، واخيرا عندما تكـون نـسبة الفقـد فـي البيانات صغيرة جدا(اقل من 5%).

أما في طريقة حذف الحالة المتوفرة فيتم التخاص من جميع المستجيبين الـذين لديهم استجابات مفقودة على كلا المتغيرين. وتقوم هذه الطريقة علـى افتـراض أن استخدام اكبر عدد من ازواج المتغيرات وأكبر كمية من البيانات المتوفرة للمستجيب يقود إلى تقديرات افضل بالنسبة للعلاقات بين ازواج المتغيرات(Witta,2002). وقد أشار جراهام(Graham,2009)، إلى افضلية طريقة حذف الحالة الكاملة في تحليـل الانحدار المتعدد أو تحليل التباين، وطريقة حذف الحالة المتوفرة في حـال ايجـاد الارتباط بين متغيرين.

2- المتوسط الحسابي للقيم: يستخدم الوسط الحسابي هذه الطريقـة(Method): القيم المفقودة إذ يتم وضع المتوسط الحسابي للسلسلة بدلا عن القيم المفقودة، ويجدر الإشارة إلى أن استخدام الوسط الحسابي هنا سيكون متأثراً بالخصائص التي يمتـاز بها الوسط لحسابي هو تأثره بالقيم المتطرفة في البيانات ويأخذ الوسـط الحـسابي الصيغة الاتية للبيانات غير المبوبة (الحيالى، 2013).

$$\bar{X} = \frac{\sum x}{n}$$

حيث أن:

$\bar{X}$:المتوسط الحسابي

$\sum x$: مجموع المشاهدات

n:عدد الأفراد

3- طريقــة حــساب قــيم تعويــضية متعــددة: Multiple Imputation (Method)(MI): وهذه الطريقة هي واحدة من الطرق التي تناولتها الدراسة الحالية، التي تعتبر من الطرق الجاذبة لمعالجة القيم المفقودة، حيث يتم فيها استبدال كل قيمة مفقودة بمتوسط مجموعه من القيم المختارة عشوائيا، ولذلك ينظر إليها على أنها تقدم قيما تعويضيـة بأخطاء معياريـة غير متحيزة فـي التحاليـل الإحصائية، وهو ما يختلف عن طريقـه حساب القيمـة التعويضية الواحدة، ويلخـص ليتـل وروبـين (Little & Rubin,2002) هذه الطريقة في ثلاث خطوات متسلـسلة هـي: خطـوة التعـويض (Imputation) ، وخطـوة التحليـل (Analysis)، وخطـوة التجميـع (Pooling).

ففي خطوة التعويض يتم حساب مجموعه من القيم التعويضية بمقدار (m) مرة، وذلك لكل قيمة من القيم المفقودة. وفي خطوة التحليل يتم إيجاد عدد من القيم التعويضية الأولية لكل قيمة مفقودة باستخدام أسلوب الانحدار. أما في خطوة التجميع فيتم إيجاد المتوسط الحسابي لجميع التقديرات الأولية للقيمة المفقودة. ويمكن جمع نتائج حسابات القيم المنفصلة بقيمة واحدة باستخدام العلاقة الاتية:

$$\overline{Q}_i = \frac{\sum_m \hat{Q}m}{M}$$

حيث أن:

$\hat{Q}$: التقدير الأولي للقيمة المفقودة.

$\overline{Q}$: المتوسط الحسابي لجميع التقديرات الأولية التعويضية، وهي التي سيتم تعويضها.

حيث يتم توفير معلومات عن دقة حساب القيمة التعويضية؛ من خلال المتوسط الحسابي والتباين. ويتكون التباين لهذه التقديرات من جزئيين:

التباين بين القيم التعويضية(B)، ويعبر عنها كما يأتي:

$$B = \frac{\sum_m (\hat{Q}m - \overline{Q}_i)^2}{M-1}$$

التباين داخل القيم التعويضي(u) ويعبر عنه كما يأتي:

$$\bar{u} = 1/m \sum \hat{u}$$

حيث أن:

u: التباين الاولي بالاعتماد على تقدير (Q).

$\bar{u}$:المتوسط الحسابي للتباين المقدر عبر القيم التعويضية (m).

كما يتم حساب التباين الكلي (T) من خلال المعادلة الاتية:

$$T = \bar{u} + \left(1 + \frac{1}{m}\right)B$$

وحساب فاعلية وكفاءة التعويض المتعدد بواسطة جداول معدة لهـذا الغـرض تتضمن نسبة الفقد وعدد مرات التعويض. وهذه الطريقة متوفرة في برنامج التحليل الإحصائي (SPSS).

تحليل الانحدار

يعتبر أداة إحصائية تقوم ببناء نموذج إحصائي وذلك لتفسير العلاقة بين متغيـر كمي واحد وهو المتغير التابع ومتغير كمي اخر أو عدة متغيرات تـسمي المـستقلة (المتنبأت)، بحيث ينتج معادلة إحصائية توضح العلاقة بـين المتغيـرات. ويمكـن استخدام هذه المعادلة في معرفة نوع العلاقة بين المتغيرات وتقدير المتغيـر التـابع باستخدام المتغيرات الاخرى (المتنبأ) وعندما تكون علاقة في النموذج الإحـصائي بين متغير تابع واحد ومتغير مستقل واحد، فإن هذا النمـوذج هـو ابـسط نمـاذج الانحدار ويسمى النموذج الخطي البسيط (simple linear regression)، وعندما تكون عدد المتغيرات المستقلة اكثر من متغير كمي واحد فإن النمـوذج يـسمى نمـوذج الانحدار المتعدد (multiple regression). ويشير (اسماعيل،2001) إلى أنه يوجـد ثلاث أهداف لاستخدام تحليل الانحدار وتتمثل: اولاً الوصف (description): حيـث يستخدم نموذج الانحدار لوصف شكل العلاقة بين المتغيـرات المـستقلة والمتغيـر التابع. ثانياً التقدير والتنبؤ(estimation and prediction): ويـتم اسـتخدام تحليـل الانحدار الخطي للقيم المستقلة للمتغير التابع المناظرة لقيمة فعلية للمتغيرات المستقلة

ويعد التقدير والتنبؤ من أهم استخدامات تحليل الانحدار في الجانب التطبيقي. واخيراً التحكم(control): ويتم من خلاله تفسير التغير في قيم المتغير التابع بدلالة التغير في قيم المتغير المستقل على أساس اتخاذ المستقل كمتغير يمكن التحكم به. وفي العـادة تستخدم طريقة المربعات الصغرى الاعتيادية(ordinary least square) لتقـدير قـيم المعلمات b،a ويستند مبدأ هذه الطريقة على تقليل مجموع مربعات المسافة بين خط الانحدار والبيانات الفعلية.

نموذج الانحدار البسيط (simple linear regression)

نموذج الانحدار البسيط هو نموذج إحصائي يقوم بتقدير العلاقة التي تربط بين متغير كمي واحد وهو المتغير التابع مع متغير كمي اخر وهو المتغير المستقل. وينتج من هذا النموذج معادلة إحصائية خطية يمكن استخدامها لتفسير العلاقة بين المتغيرين أو تقدير قيمة المتغير التابع عند معرفة قيمة المتغير المستقل ويمكن صياغة العلاقة الإحصائية بالنموذج التالي:

$$\hat{y}_i = \hat{a} + b x_i + \hat{e}$$

حيث أن:

$\hat{y}_i$: هو المتغير التابع

x_i: هو المتغير المستقل

$\hat{e}$: هو الخطأ العشوائي

$\hat{a}$: هي قيمة ثابتة تعبر عن قيمة $\hat{y}_i$ عندما تكون قيمة x_i تساوي الصفر

b: تعبر عن الخط المستقيم الذي يوضح العلاقة.

ويفترض النموذج أن $\hat{y}_i$ متغير عشوائي يتبع التوزيع الطبيعي بمتوسط يرتبط بقيمة x_i وبتباين ثابت على اختلاف قيم x_i. ويمكن تقدير معالم النموذج باستخدام طريقة المربعات الصغرى.

غير أنه ينبغي فهم أن الانحدار يتأثر بأي المتغيرين يتم وضعه على المحور الافقي (x)، وأيهما يوضع على المحور الراسي (y). إذ تنشأ أشكالا مختلفة إلى حد ما في الانحدار، اعتمادا على اختيارك(علام، 2005).

ويشير (ربيع، 2008) للحكم على صلاحية نموذج الانحدار الذي تم توفيقة

للعلاقة بين المتغير التابع والمتغيرات التفسيرية، لا بد وأن يتوافر مجموعة من الشروط يمكن تقسيمها إلى.

أولا: الشرط النظري:

في تحليل الانحدار يجب ان يتم التحقق من بعض الافتراضات حتى نقوم بالتحليل وعند حدوث خال في تلك الفروض فانه يتم التوقف لعدم استيفاء الشروط اللازمة لإتمام عملية تحليل الانحدار وهي افتراضات مهمه يجب توفرها باختبارها احصائيا.

قبول القدرة التفسيرية للنموذج:

يقصد بالقدرة التفسيرية لنموذج الانحدار: مدى قدرة المتغيرات المستقلة في النموذج على تفسير التغيرات التي تحدث في المتغير التابع وتعزى إلى المتغيرات المستقلة(ربيع،2008).

ثانيا: الشروط الرياضية:

تتضمن المعنوية الكلية لنموذج الانحدار :

يقصد بها اختيار الشكل الدالي للعلاقة بين المتغير التابع والمتغيرات التفسيرية في نموذج الانحدار . وذلك باستخدام اختبار (F-test). فمثلا، لو أننا قد اخترنا نموذج الانحدار الخطي لتمثل العلاقة بين المتغير التابع والمتغيرات المستقلة، فإن اختيار المعنوية الكلية: يقصد به الاجابة على السؤال التالي: هل الشكل الدالي المقترح (النموذج الخطي) هو نموذج مقبول لتمثيل العلاقة بين المتغير التابع والمتغيرات التفسيرية ام لا؟. بحيث في حالة النفي فإنه على الباحث محاولة ايجاد نموذج اخر يمكن أن يقدم وصف افضل للعلاقة بين متغيرات النموذج، كان يقترح نموذج غير خطي لهذه العلاقة.

كما أن قبول الشكل الدالي المقترح له مدلول اخر: وهو أن هناك معامل واحد على الاقل من معاملات نموذج الانحدار يختلف عن الصفر(معنوي).

المعنوية الجزئية للنموذج:

يقصد بها معنوية معاملات الانحدار لكل متغير من المتغيرات التفسيرية على حده بالإضافة إلى ثابت الانحدار. وذلك من خلال اختبار (t-test).

مدى توافر الشروط للطريقة المستخدمة في تقدير معالم نموذج الانحدار.

يعتبر اشهر هذه الطرق: طريقة المربعات الصغرى العادية OLS. وتتمثل أهـم شروط هذه الطريقة في:

أ- **اعتدالية التوزيع الاحتمالي للبواقي (Normality test) :**

لكي يمكن استخدام كل من اختبار (f) واختبار (t) ، سواء عند اختبار المعنوية الكلية أو المعنوية الجزئية فلنموذج الانحدار، يلزم توافر شرط اعتدآليـة التوزيـع الاحتمالي للبواقي.

ونود الإشارة إلى أن التقيد بهذا الشرط مرتبط بحجم العينة، اذ يعتبـر شـرطا ضروريا في حالة العينات الصغيرة، أما في حالات العينات الكبيرة فيمكن التخلـي عنه، وذلك لأنه وفقا لنظرية الحد المركزية نجد أن التوزيعات الاحتمالية تؤول إلى التوزيع الطبيعي فـي حالـة العينـات التـي تزيـد حجمهـا عـن 30 مـشاهده (Palta & Mari ,2003).

ب- الاستقلال الذاتي للبواقي:

ترجع أهمية دراسة الارتباط الذاتي للبواقي في تحليل الانحدار، إلى أن وجـود هذا الارتباط من شأنه أن يجعل قيمة التباين المقدر للخطأ يكون بأقل مـن قيمتـه الحقيقة. وبالتالي فإن قيمة احصاءات الاختبار التي تعتمد على هذ التباين مثـل(t)، (f)، (r^2) تكون اكبر من قيمتها الحقيقة، مما يجعل القرار الخاص بجودة توفيـق النموذج قرار مشكوك في صحته. وبشكل عام، ليس هناك حد فاصل متفق عليه لهذه النسبة. بل انه امر نسبي وتقديري يتوقف على طبيعة الظاهرة التـي تحكـم هـذه العلاقة. مثال على ذلك، لو اننا نريد توفيق نموذج انحدار فـي احـدى الكـوارث الطبيعية كالزلازل، فانه في هذه الحالة لو أن القدرة التفسيرية للنمـوذج المقتـرح تتراوح بين 30% الى 40% في هذه الحالة نقول انه نموذج جيد على العكس مـن ذلك قد تكون النسبة 60% في ظاهرة اخرى كالظاهرة الاقتصادية نجدها غير كافية (ربيع، 2008).

ج- اختبار تجانس البواقي (اختبار ثبات التباين) Homoscedasticity:

إن عدم ثبات التباين في نموذج الانحدار من شأنه أن يترتب عليه نفس الاثار المترتبة في حالة وجود ارتباط ذاتي بين البواقي، حيث تكون الاخطاء المعيارية مقدرة باقل من قيمتها الحقيقية. وبالتالي تصبح هذه التقديرات متحيزة biased، الأمر الذي يجعل نتائج الاستدلال الإحصائي مشكوك في صحتها (Berk ,Richard 2003). وهذا معناه يجب أن يكون هناك ثبات التغير (التباين) في قيمة البواقي. عندما يكون هناك تجانس فإن البواقي ستكون متساوية إلى حد ما عند جميع القيم أو بمعنى آخر لن نلاحظ اتجاه لزيادة أو نقصان البواقي مع تغير قيمة المتغير المستقل.

3.2 الدراسات السابقة:

تبحث هذه الدراسة في موضوعين رئيسين هما: أثر نسبة القيم المفقودة، وطريقة معالجتها على دقة تقدير معالم معادلة الانحدار البسيط . الأمر الذي تطلب الرجـوع إلى جميع الدراسات – ما امكن – المتصلة بهما، ومن خلال اطلاع الباحـث علـى الدراسات السابقة المتصلة بموضوع تعويض القيم المفقودة، تبين للباحث وجود عدد من الدراسات تناولت طرق تعويض القيم المفقودة على سمات مختلفة لكنها لم تتناول بشكل مباشر أثرها على معالم معادلة الانحدار البسيط أو المتعدد، وعدد قليـل مـن الدراسات الخاصة بنسب الفقد .

ومن هذه الدراسات دراسة وايمان(Wayman , 2003) طبق الباحـث ثـلاث طرق لمعالجة القيم المفقودة في البيانات وتمت المقارنة بين هذه الطرق لمعرفـة اي الطرق تعطي تحيز اقل من غيرها وهذه الطرق هي: طريقة حساب قيم تعويـضية متعددة وطريقة حذف الحالة وطريقة تعويض الوسط الحسابي، قامت هذه الدراسـة على ثلاث متغيرات هي: الجنس (ذكور، إناث) نوع التعلـيم (حكـومي، خـاص) الصف الدراسي(سادس، سابع، ثامن) وكانت نسب الفقـد بـين(11% إلـى 18%) وكانت عينة الدراسة مكونة من 19373 طالب وطالبة من طلبـة المـدارس فـي الولايات المتحدة وكانت نسبة الفقد 15% من العدد الكلي للاستجابات، وبينت نتائج الدراسة إلى أن طريقة حساب قيم تعويضية متعددة اعطت نتائج ذات تحيز اقل من

حيث الاخطاء المعيارية مقارنة مع بقية الطرق المستخدمة في الدراسة .

وفي دراسة لهوثورن و إليوت (Hawthorn & Eliot , 2005) والتي هدفت إلى تحديد الطريقة الأنسب لمعالجة البيانات المفقودة، حيث تم المقارنة بين الطرق التالية لمعالجة القيم المفقودة: طريقة الوسط الحسابي للفرد، حذف الحالة، حساب قيمة تعويضية من توزيعات غير مشروطة، والوسط الحسابي للفقرة، حساب قيمة تعويضية من خلال الانحدار. وكانت عينة الدراسة مكونه من (1200) فرد، وطبق الباحثان خمسة احجام مختلفة من العينة كالتالي (25،50،100،200،400) على هذه الفقرات، وظهرت نسب متفاوتة من القيم المفقودة كالتالي (20%،40%،60%) ليتم مقارنة مجموعتين من العينات الأولي تفترض عدم وجود قيم مفقودة فيها والثانية بافتراض وجود قيم مفقودة على النحو التالي (20%،40%،60%) وذلك باستخدام اختبار (t-test) لمقارنة مجموعتين من العينات، خلصت الدراسة أن هناك افضلية لطريقتي المتوسط الحسابي للفرد وطريقة حساب قيم تعويضية من توزيعات غير مشروطة بالمقارنة مع بقية طرق المعالجة للبيانات.

واجرى أليسون (Alison ,2006) دراسة محاكاة لمعرفة أثر استخدام طرق حساب قيم تعويضية مختلفة لتعويض القيم المفقودة على البيانات التصنيفية، حيث قام الباحث بتوليد بيانات لعينة مكونة من (500) فرد، وبنسب فقد مختلفة هي: (1%،5%،20%،50%) بحيث استخدم آلية الفقد العشوائي بالكامل، وآلية الفقد عشوائياً، ولتحقيق أهداف الدراسة استخدم خمسة طرق لمعالجة البيانات المفقودة وهي: طريقة حساب قيمة تعويضية من خلال دالة الاستجابة، طريقة حساب قيمة تعويضية من خلال الانحدار اللوجستي وطريقة الحالة الكاملة، وطريقة حساب قيمة تعويضية خطية بدون تقريب، وطريقة حساب قيمة تعويضية مع التقريب، وخلصت نتائج الدراسة بناءاً على الأوساط الحسابية للأخطاء المعيارية والأوساط الحسابية للانحرافات المعيارية في كل من أليتي الفقد؛ الفقد العشوائي بالكامل والفقد العشوائي إلى أن الوسط الحسابي المقدر يتعرض للتحيز إذا تم تدوير القيم التعويضية المقدرة إلى اعداد صحيحة (0 أو1). وأنه لا يتعرض الوسط الحسابي المقدر للتحيز، إذا لم يتم تدوير القيم المقدرة إلى اعداد صحيحة. وأن طريقة "حساب قيم تعويضية مع

التقريب" تعطي نتائج متحيزة وهي دون مستوى الطرق الاخرى. واخيـرا يفـضل استخدام آلية الفقد العشوائي بدلا من آلية الفقد العشوائي بالكامل لامتلاكهـا اخطـاء معيارية اقل.

وفي دراسة (بني عواد، 2010) التي هدفت لمقارنة طرق التعامل مع البيانـات المفقودة في تقدير معالم الفقرات وقدرات الأفراد حيث تمت معالجة البيانات المفقودة باستخدام ثمان طرق مختلفة من طرق معالجة القيم المفقودة باستخدام طرق مختلفـة من طرق معالجة القيم المفقودة، وتم تطبيق اختبار أوتيس- لينون للقدرة العقلية على عينة مكونة من 1600طالب من طلبة الصف الثامن في مدارس تربية اربد الأولـى موزعين على 20 مدرسة تم اختيارها عشوائيا خلال العام الدراسي 2009-2010، وقد خلصت النتائج إلى وجود فروق ذات دلالة إحصائية بين المتوسطات الحـسابية للأخطاء المعيارية لتقدير معلمة القدرة للأفراد، وكذلك في معالم الفقرات (الصعوبة، التميز، التخمين) تعزى لطريقة المعالجة، وطول الاختبار، والتفاعل بينهمـا حيـث كانت التقديرات الكثر دقة عند طريقة دالة الاستجابة وطول الاختبار (80) فقرة .

اجرى كوكلاك وكيري(Cokluk & Kayri,2011) دراسة استخدم فيها خمسة من طرق معالجة البيانات بنسب فقد مختلفة، تم تقسيم نسب الفقد إلى فئتـين: الأولـى تراوح نسب الفقد بين(15%) و (20%) والثانية تراوح نسب الفقد فيها بين(0%) و (15%)، ومن ثم تمت مقارنة ثلاثة معاملات ثبات ناتجة من تطبيق هـذه الطـرق وهي: معامل الارتباط المصحح، ومعامل كرونباخ الفا للاتساق الداخلي، والبنـاءات العامليه، وكانت عينة الدراسة مكونة من 200 معلم من المرشحين لوظيفة مـدرس والملتحقين بقسم التعليم الابتدائي بكلية العلوم التربوية في جامعة أنقرة للفصل الثاني للعام الدراسي 2009/2008. وقد خلصت النتائج إلى أن الطرق المختلفة في معالجة القيم المفقودة تتسبب في تقليل نسب التباين المفسر للطرق المستخدمة في الدراسـة. كما اظهرت النتائج أن معاملات ثبات كرونباخ الفا للاتساق الـدا خلـي والجـذور الكامنة في المجموعة الأولى التي تحوي فيها قيم مفقودة تراوح نسبتها بين (0% - 15%)، وكذلك المجموعة الثانية التي تحوي قيم مفقودة تتراوح نسبتها بين (20%- 15%)، قد انخفضت بسبب تعويض القيم المفقودة في البيانات.

اجرى (الدرابسة، 2012) دراسة هدفت إلى بيان أثر طريقة معالجة القيم المفقودة، وطريقة تقدير قدرات الأفراد على دقة تقدير معالم الفقرات والأفراد. ولتحقيق أهداف الدراسة، تم استخدام بيانات مولدة لاستجابات 1500 مستجيب على اختبار مكون من 80 فقرة ثنائية الاستجابة ومطابقه للنموذج ثلاثي المعلمة، ليتم الحصول على بيانات تتضمن استجابات مفقودة بنسبة(5%)، وتمت معالجة القيم المفقودة بثلاث طرق هي: تعظيم التوقعات (EM)، والقيم التعويضية المتعددة (MI)، ودالة الاستجابة(RF). وتم تقدير معالم الفقرات والأفراد بطريقتي الارجحية العظمى (ML)، وبييز للتوقع البعدي(EAP)، حيث اظهرت النتائج وجود اختلاف في دقة تقدير معلمة التميز يعزى لاختلاف طريقة التقدير ولصالح طريقة الارجحية العظمى(ML)، ووجود اختلاف في دقة تقدير معلمة الصعوبة يعزى لاختلاف طريقة التقدير ولصالح طريقة بييز للتوقع البعدي(EAP). إن وجود اختلاف في دقة تقدير معلمة التميز يعزى لاختلاف طريقة المعالجة للقيم المفقودة، ولصالح طريقة القيم المتعددة(MI)، ووجود اختلاف يعزى للتفاعل بين متغيري طريقة التقدير وطريقة المعالجة. وعدم وجود اختلاف في دقة تقدير معلمة الصعوبة أو معلمة التخمين يعزى لاختلاف طريقة المعالجة للقيم المفقودة أو طريقة التقدير أو للتفاعل بينهما. ووجود اختلاف في دقة تقدير قدرات الأفراد يعزى لطريقة معالجة القيم المفقودة ولصالح طريقة تعظيم التوقعات(EM)، ووجود اختلاف في دقة تقدير قدرات الأفراد يعزى لطريقة التقدير المستخدمة ولصالح طريقة بييز للتوقع البعدي(EAP). ويوجد اختلاف في دقة تقديرات قدرات الأفراد يعزى للتفاعل بين متغيري طريقة المعالجة للقيم المفقودة وطريقة التقدير.

كما اجرى (الزعبي، 2013) دراسة هدفت إلى بيان أثر نسبة البيانات المفقودة وطريقة التعويض عنها في دقة تقدير معالم الفقرات والأفراد. ولتحقيق هذه الأهداف، تم استخدام بيانات مولدة لتوليد استجابات (1400) مستجيب على اختبار مكون من (100) فقرة ثنائية الاستجابة، حيث تراوحت قيم معلمة التميز لفقرات الاختبار بين (0.1) و (2)، والصعوبة بين (2.5-) و (2.5) بافتراض أن قدرات المستجيب تتوزع توزيعا طبيعيا وباستخدام برنامج (SPSS) و(EXCEL)، تم

الحصول على بيانات تتضمن بيانـات مفقـودة بنـسبة (5%،15%،20%،30%)، وتمت معالجة هذه الاستجابات بطريقتي المعالجة للقيم المفقودة؛ تعظـيم التوقعـات، والقيم التعويضية المتعددة، حيث اظهرت النتائج وجود أثر دال إحصائيا فـي دقـة تقدير معلمة صعوبة الفقرات يعزى لكل من: متغير طريقة التعويض في النمـوذج الاحادي المعلمة ولصالح طريقة حساب قيم تعويضية متعددة. ومتغير نسبة الفقد في النموذجين الاحادي المعلمة والثنائي المعلمة، بأفضلية بنسبة فقد (5%) .

كما اظهرت النتائج وجود أثر دال إحصائيا في دقة تقدير معلمة تميز الفقـرات في النموذج الثنائي المعلمة يعزى لكل من: متغير طريقة التعويض ولصالح طريقـة حساب قيم تعويضية متعددة. ومتغير نسبة الفقد بأفضلية نسبة فقد(5%).

4.2 التعقيب على الدراسات السابقة:

في ضوء ما تقدم من دراسات حول طرق تعويض القيم المفقودة، تبين أن بعض هذه الدراسات قد تناولت طرق تعويض القيم المفقودة، وأثرها على دقة تقدير معالم الفقرات وقدرات المستجيبين في ضوء نظرية استجابة الفقرة، وبعضها تناولت أثرها على صدق وثبات المقاييس، وبعضها تناولت أثرها على التحيز في تقدير المعالم ولا توجد دراسات في حدود علم الباحث تناولت أثر طـرق تعـويض القـيم المفقـودة باختلاف نسب الفقد على دقة تقدير معالم معادلة الانحدار لذلك جاءت هذه الدراسـة لتحقيق هذا الهدف وذلك باستخدام بيانات مولدة لمعرفة أثر نسب الفقد في البيانـات وطرق تعويضها على دقة تقدير معالم معادلة الانحدار البسيط .

الفصل الثالث
المنهجية والتصميم

يتناول هذا الفصل وصفا للإجراءات التي تم إتباعها لتوليد البيانات، وآلية فقد البيانات بنسب مختلفة ومعالجتها بالطرق الإحصائية التي بنيت عليهـا الدراسـة، والتحقق من إفتراضات معادلة الانحدار البسيط، وأساليب التحليل الإحصائي التي تم استخدامها للإجابة عن الدراسة.

1.3 منهجية الدراسة:

تم إجراء بحث اساسي يهدف إلى استقصاء أثر نسبة القـيم المفقـودة وطـرق معالجتها على دقة تقدير معالم معادلة الانحدار البسيط ويعتمد استخدام منهج المحاكاة التجريبي على المستجيبين ، لتحديد الافضلية في طرق المعالجة المستخدمة وتـأثير نسب الفقد المستخدمة في الدراسة.

2.3 إجراءات الدراسة:

1.2.3 توليد البيانات:

وللإجابة عن اسئلة الدراسة تم اتباع الإجراءات الاتية:

تم توليد البيانات باستخدام برنامج (WINGEN3)، حيث تم توليد اختبارين طول كل منها 12 فقرة، وتم توليد توزيع قدرات(200) مستجيب وهـي تمثـل مجتمـع الدراسة. بتوزيع طبيعي لقدراتهم بمتوسط(صفر) وانحرافه المعيـاري(1) اعتمـادا على إشارة لورد(Lord,1980)، بأن هذه الاعداد تمثل الحد الادنى للحـصول علـى افضل التقديرات. وتم ذلك وفقا للخطوات التالية:

تم توليد البيانات الاصـلية للحـصول علـى متغيـرين(x)المتغيـر المـستقل (independent) والمتغير (y) المتغير التابع (dependent).

تم عمل تحليل انحدار بسيط للبيانات الاصلية لمعرفة قيمة معـاملات الانحـدار ومدى تحقق افتراضات تحليل الانحدار البسيط.

تم فقد القيم بطريقة عشوائية بنسب مختلفة (5%،10%،15%،20%) باستخدام برنامج الإحصائية للعلوم الاجتماعية (SPSS) وفقا للخطوات التالية:

عمل فقد بنسبة 5% من عدد استجابات الأفراد على الفقـرات فـي مـصفوفة المتغير المستقل باستخدام البرنامج الإحصائي (SPSS).

عمل فقد بنسبة 10% من عدد استجابات الأفراد على الفقرات فـي مـصفوفة المتغير المستقل باستخدام البرنامج الإحصائي (SPSS).

عمل فاقد بنسبة 15% من عدد استجابات الأفراد على الفقرات فـي مـصفوفة المتغير المستقل باستخدام البرنامج الإحصائي (SPSS).

عمل فاقد بنسبة 20% من عدد استجابات الأفراد على الفقرات فـي مـصفوفة المتغير المستقل باستخدام البرنامج الإحصائي (SPSS).

تم تعويض القيم المفقودة بثلاث طرق (الحذف، الوسـط الحـسابي، التعـويض المتعدد). باستخدام برنامج (SPSS) وفقا للخطوات التالية:

أولاً: تم معالجة القيم المفقودة بنسب(5%،10%،15%،20%) بطريقة الحـذف كلا على حدة وعمل تحليل انحدار بسيط لكل نسبة لمعرفة مـدى التـأثير على معاملات الانحدار وافتراضاته الواجب توفرها مـن حيـث(معامـل الارتباط المعدل، والمعنوية الجزئية لنموذج الانحدار البسيط قيمة اختبـار t، والمعنوية الكلية لنموذج الانحدار البسيط قيمة اختبار f.

ثانياً: تم معالجة القيم المفقودة بنسب(5%،10%،15%،20%) بطريقة الوسـط الحسابي كلا على حدة وعمل تحليل انحدار بسيط لكل نسبة لمعرفة مـدى التأثير على معـاملات الانحـدار وافتراضـاته الواجـب توفرهـا مـن حيث(معامل الارتباط المعدل، والمعنوية الجزئية لنموذج الانحدار البـسيط قيمة اختبار(t)، والمعنوية الكلية لنموذج الانحدار البسيط قيمة اختبار (f).

ثالثاً: تم معالجة القيم المفقودة بنسب(5%،10%،15%،20%) بطريقة التعويض المتعدد كلا على حدة وعمل تحليل انحدار بسيط لكل نسبة لمعرفة مـدى التأثير على معـاملات الانحـدار وافتراضـاته الواجـب توفرهـا مـن حيث(معامل الارتباط المعدل، والمعنوية الجزئية لنموذج الانحدار البـسيط

قيمة اختبار (t)، والمعنوية الكلية لنموذج الانحدار البسيط قيمة اختبار (f).

2.2.3 فقد القيم:

تم فقد القيم بطريقة عشوائية بنسب مختلفة (5%،10%،15%،20%) باستخدام البرنامج الإحصائي (SPSS) وفقا للخطوات التالية:

عمل فقد بنسبة 5% من عدد استجابات الأفراد على الفقرات في مصفوفة المتغير المستقل باستخدام البرنامج الإحصائي (SPSS).

عمل فقد بنسبة 10% من عدد استجابات الأفراد على الفقرات في مصفوفة المتغير المستقل باستخدام البرنامج الإحصائي (SPSS).

عمل فاقد بنسبة 15% من عدد استجابات الأفراد على الفقرات في مصفوفة المتغير المستقل باستخدام البرنامج الإحصائي (SPSS).

عمل فاقد بنسبة 20% من عدد استجابات الأفراد على الفقرات في مصفوفة المتغير المستقل باستخدام البرنامج الإحصائي (SPSS).

3.2.3 معالجة القيم المفقودة:

تم تعويض القيم المفقودة بثلاث طرق (الحذف، الوسط الحسابي، التعويض المتعدد). باستخدام برنامج (SPSS) وفقا للخطوات التالية:

أولاً: تم معالجة القيم المفقودة بنسب(5%،10%،15%،20%) بطريقة الحذف كلا على حدة وعمل تحليل انحدار بسيط لكل نسبة لمعرفة مدى التأثير على معاملات الانحدار وافتراضاته الواجب توفرها من حيث(معامل الارتباط المعدل، والمعنوية الجزئية لنموذج الانحدار البسيط قيمة اختبار t ، والمعنوية الكلية لنموذج الانحدار البسيط قيمة اختبار f.

ثانياً: تم معالجة القيم المفقودة بنسب(5%،10%،15%،20%) بطريقة الوسط الحسابي كلا على حدة وعمل تحليل انحدار بسيط لكل نسبة لمعرفة مدى التأثير على معاملات الانحدار وافتراضاته الواجب توفرها من حيث(معامل الارتباط المعدل، والمعنوية الجزئية لنموذج الانحدار البسيط قيمة اختبار(t)،

والمعنوية الكلية لنموذج الانحدار البسيط قيمة اختبار (f).

ثالثاً: تم معالجة القيم المفقودة بنسب(5%،10%،15%،20%) بطريقـة التعـويض المتعدد كلا على حدة وعمل تحليل انحدار بسيط لكل نسبة لمعرفة مدى التأثير على معاملات الانحدار وافتراضاته الواجب توفرها من حيث(معامل الارتباط المعدل، والمعنوية الجزئية لنموذج الانحدار البسيط قيمة اختبار (t)، والمعنوية الكلية لنموذج الانحدار البسيط قيمة اختبار (f).

4.2.3 تقدير معاملات الانحدار البسيط للقيم

بعد أن يتم جمع درجات الأفراد على الفقرات للمتغير التـابع(Y) والمتغيـر المستقل(X) والتي تعتبر البيانات الحقيقية بدون أن يكون هناك اي قيمة مفقودة. تـم استخراج معاملات الانحدار بعد الفقد بالنسب المختلفة ومعالجتها بالطرق المـذكورة وعمل مقارنه بينها ومن ثم الحكم على افضل طرق العلاج وتأثير كل نسبة فقد على دقة معاملات الانحدار البسيط.

الفصل الرابع

عرض النتائج ومناقشتها والتوصيات

يتضمن هذا الفصل عرضا للنتائج التي تم التوصل إليها من خلال هذه الدراسة، والتي هدفت إلى فحص أثر نسبة القيم المفقودة وطرق معالجتها في دقة تقدير معالم معادلة الانحدار البسيط باستخدام بيانات مولدة؛ بالإضافة إلى مناقشة هـذه النتـائج وتفسيرها، وقد تم عرض هذه النتائج بحسب اسئلة الدراسة ومتغيراتها حسب منهجية منظمة تقوم على عرض السؤال ثم نوع الإحصائي المناسب، يلـي ذلـك جـدول البيانات ثم التعليق عليها.

1.4 عرض النتائج ومناقشتها

النتائج المتعلقة بالسؤال الأول الذي ينص على: " ما أثر طرق معالجة البيانـات المفقودة (الحذف، حساب المتوسط، التعويض المتعدد).على دقة معـاملات معادلـة الانحدار البسيط ؟".

للإجابة عن السؤال الأول، تم حساب قيمة معاملات الانحدار البـسيط الثابـت والميل وقيمة (t-test) والمعنوية لكل معامل من معاملات الانحدار البـسيط حـسب طريقة المعالجة و الجدول (1) يبين ذلك.

الجداول (4،3،2،1): معاملات الانحدار وقيمة (t-test) والمعنوية لكـل معامـل وفقا لمتغيري نسبة الفقد في البيانات، وطريقة معالجتها بالحذف.

جدول (1)

أثر طريقة المعالجة بالحذف على معاملات الانحدار البسيط

طريقة المعالجة	نسبة الفقد	معامل الارتباط المصحح	معامل Durbin	معاملات الانحدار البسيط A	معاملات الانحدار البسيط B	قيمة t	المعنوية
البيانات الاصلية	0	0.96	1.991	4.605	0.860	5.402	. 000
الحذف	%5	0.94	1.991	5.150	0.854	5.159	.000
	%10	0.91	1.997	7.481	0.827	5.970	.000
	%15	0.83	2.237	12.28	0.769	7.257	.000
	%20	0.74	2.139	20.70	0.656	10.91	.000

يلاحظ من بيانات الجدول (1) أن قيمة معامل الارتباط المصحح تقل مع زيادة نسبة الفقد بينما معامل دوربن يزداد بزيادة نسبة الفقد وايضا معامل الانحدار الثابت(a) يزداد على العكس من معامل الانحدار(b) الميل فان قيمته تقل مع زيادة نسبة الفقد في البيانات أما الإحصائي (t) فإن قيمته تزداد بزيادة نسبة الفقد ومع ذلك نجد ان جميع هذه المعاملات تدل على تحقق فرض المعنوية الجزئية لنموذج الانحدار البسيط مع وجود فروق ظاهرية بين معاملات الانحدار للثابت (a) في معادلة الانحدار البسيط وفقا لمتغيري؛ نسبة الفقد في البيانات وطرق معالجتها. ولمعرفة الدلالة الإحصائية لتلك الفروق الظاهرية وفقا لمتغيري نسبة الفقد في البيانات وطرق معالجتها ؛ تم استخدام تحليل الانحدار البسيط للتحقق من الافتراضات الواجب توفرها في نموذج الانحدار البسيط.

جدول (2)

أثر طريقة المعالجة بالوسط الحسابي على معاملات الانحدار البسيط

طريقة المعالجة	نسبة الفقد	معامل الارتباط المصحح	معامل Durbin	معاملات الانحدار البسيط A	B	قيمة t	المعنوية
البيانات الاصلية	0	0.96	1.991	4.605	0.860	5.402	. 000
	%5	0.96	1.992	4.580	0.860	5.338	.000
الوسط الحسابي	%10	0.96	2.004	4.381	0.863	5.046	.000
	%15	0.96	2.081	3.547	0.875	3.787	.000
	%20	0.95	2.093	2.541	0.889	2.573	.011

يلاحظ من بيانات الجدول (2) إن قيمة معامل الارتباط المصحح ثابتة حيث بغلت بين(0.95-0.96) وكذلك معامل دوربن يزداد بنسبة قليلة بزيادة نسبة الفقد وايضا معامل الانحدار(a) الثابت يقل بزيادة نسبة الفقد والمعالجة بالوسط الحسابي في حين نجد ان معامل الانحدار(b) الميل تزداد قيمته مع زيادة نسبة الفقد في البيانات أما الإحصائي (t) فإن قيمته تقل بزيادة نسبة الفقد ومع ذلك نجد أن جميع

هذه المعاملات تدل على تحقق فرض المعنوية الجزئية لنموذج الانحدار البسيط مع وجود فروق ظاهرية بين معاملات الانحدار للثابت (a) في معادلة الانحدار البسيط وفقا لمتغيري؛ نسبة الفقد في البيانات وطرق معالجتها. ولمعرفة الدلالة الإحـصائية لتلك الفروق الظاهرية وفقا لمتغيري نسبة الفقد في البيانات وطرق معالجتهـا ؛ تـم استخدام تحليل الانحدار البسيط للتحقق من الافتراضات الواجب توفرها في نمـوذج الانحدار البسيط.

جدول (3)

أثر طريقة المعالجة بالتعويض المتعدد على معاملات الانحدار البسيط

طريقة المعالجة	نسبة الفقد	معامل الارتباط المصحح	معامل Durbin	معاملات الانحدار البسيط A	B	قيمة t	المعنوية
التعويض المتعدد	%5	0.96	1.970	4.628	0.860	5.471	.000
	%10	.096	1.947	4.712	0.858	5.506	.000
	%15	0.96	1.990	4.885	0.859	5.379	.000
	%20	0.95	1.892	5.240	0.853	5.442	.000

يلاحظ من بيانات الجدول (3) أن قيمة معامل الارتباط المصحح ثابت تتراوح قيمته بين(0.95-0.96) وكذلك معامل دوربن تقل قيمته بنسبة قليلة جداً بزيادة نسبة الفقد وايضا معامل الانحدار(a) الثابت يزداد على العكس من معامـل الانحـدار(b) الميل فان قيمته ثابتة تقريباً مع زيادة نسبة الفقد في البيانات أما الإحصائي (t) فـإن قيمته قريبة جداً في حالات الفقد والمعالجة بالتعويض المتعدد من القيمة الاصـلية، ومع ذلك نجد أن جميع هذه المعاملات تدل على تحقق فـرض المعنويـة الجزئيـة لنموذج الانحدار البسيط مع وجود فروق ظاهرية بسيطة بين معـاملات الانحـدار للثابت (a) في معادلة الانحدار البسيط وفقا لمتغيري؛ نسبة الفقد في البيانات وطرق معالجتها. ولمعرفة الدلالة الإحصائية لتلك الفروق الظاهرية وفقا لمتغيري نسبة الفقد في البيانات وطرق معالجتها؛ تم استخدام تحليل الانحـدار البـسيط للتحقـق مـن

الافتراضات الواجب توفرها في نموذج الانحدار البسيط.

وللحكم على معنوية النموذج ككل نستخدم تحليل التباين وفقا للجدول رقم (6،5،4) الذي يبن مستوى الدلالة العلمية حسب متغيرات الدراسة ؛ نسبة الفقد في البيانـات، وطرق معالجتها والجداول (6،5،4) تبين ذلك.

جدول (4)

جدول تحليل التباين لطريقة المعالجة بالحذف وفقا لنسب الفقد(5%،10%،15%،20%)

مصدر التباين	مجموع المربعات	درجة الحرية	متوسط المربعات	قيمةF	المعنوية
*حقيقي					
الانحدار	15786.597	1	15786.597	5112.416	.000
الخطأ	611.403	198	3.088		
المجموع	16398.000	199			
*فقد 5%					
الانحدار	15556.766	1	15556.766	3661.571	.000
الخطأ	841.234	198	4.249		
المجموع	16398.000	199			
فقد10%					
الانحدار	15013.761	1	15013.761	2147.551	.000
الخطأ	1384.239	198	6.991		
المجموع	16398.000	199			
فقد15 %					
الانحدار	13667.133	1	13667.133	990.928	.000
الخطأ	2730.867	198	13.792		
المجموع	16398.000	199			
فقد 20%					
الانحدار	12234.970	1	12234.970	560.763	.000
الخطأ	4320.050	198	21.818		
المجموع	16555.020	199			

الجدول(4) يبين المعنوية الكلية لنموذج الانحدار البسيط للبيانات الاصلية والبيانات بنسب الفقد المختلفة(5%،10%،15%،20%) والتي تم معالجتها بطريقة الحذف بتحليل التباين.

يتضح من جدول تحليل التباين(4): أن قيمة الاحتمال P.value تساوي صفر، وهي اقل من مستوى المعنوية 5%، وبالتالي نرفض الفرض الصفري القائل بأن نموذج الانحدار غير معنوي، وهذا يعني أن هناك واحد على الاقل من معاملات الانحدار تختلف عن الصفر. للحكم على معنوية نموذج تحليل الانحدار البسيط وفقا لطريقة المعالجة بالوسط الحسابي بنسب فقد للبيانات(5%،10،15%،20%) .

الجدول (5)

جدول تحليل التباين لطريقة المعالجة بالوسط الحسابي وفقا لنسب الفقد(5%،10%،15%،20%)

مصدر التباين	مجموع المربعات	درجة الحرية	متوسط المربعات	قيمةF	المعنوية
معالجةبالوسط5%					
الانحدار	15779.375	1	15779.375	5050.4	.000
الخطأ	618.625	198	3.124		
المجموع	16398.000	199			
معالجةبالوسط10%					
الانحدار	15768.949	1	15768.949	4963.4	.000
الخطأ	629.051	198	3.177		
المجموع	16398.000	199			
معالجةبالوسط15%					
الانحدار	15688.621	1	15688.621	4378.9	.000
الخطأ	709.379	198	3.583		
المجموع	16398.000	199			
معالجةبالوسط20%					
الانحدار	15787.767	1	15787.767	4074.2	.000
الخطأ	767.253	198	3.875		
المجموع	1655.020	199			

يتضح من جدول تحليل التباين(5): أن قيمة الاحتمال P.value تساوي صفر، وهي اقل من مستوى المعنوية 5%، وبالتالي نرفض الفرض الصفري القائل بأن نموذج الانحدار غير معنوي، وهذا يعني أن هناك واحد على الاقل من معاملات الانحدار تختلف عن الصفر.

للحكم على معنوية نموذج تحليل الانحدار البـسيط وفقـا لطريقـة المعالجـة بالتعويض المتعدد بنسب فقد للبيانات(5%،10،15%،20%) والجـدول رقـم (6) يبين ذلك .

جدول (6)

جدول تحليل التباين لطريقة المعالجة بالتعويض المتعدد وفقا لنسب الفقد(5%،10%،15%،20%)

معالجة التعويض	مصادر البيانات	مجموع المربعات	درجة الحرية	متوسط المربعات	قيمة F	المعنوية
%5	الانحدار	15783.52	1	15783.525	5085.87	.000
	الخطأ	614.475	198	3.103		
	المجموع	16398.00	199			
%10	الانحدار	15779.78	1	15779.782	5053.88	.000
	الخطأ	618.218	198	3.122		
	المجموع	16398.00	199			
%15	الانحدار	15657.07	1	15657.067	4184.05	.000
	الخطأ	740.933	198	3.742		
	المجموع	16398.00	199			
%20	الانحدار	15760.42	1	15760.420	3927.21	.000
	الخطأ	794.600	198	4.013		
	المجموع	16555.02	199			

يتضح من جدول تحليل التباين(6): أن قيمة الاحتمال P.value تساوي صفر، وهي اقل من مستوى المعنوية 5%، وبالتالي نرفض الفرض الصفري القائـل بـأن

نموذج الانحدار غير معنوي، وهذا يعني أن هناك واحد على الاقل مـن معـاملات الانحدار تختلف عن الصفر.

النتائج المتعلقة بالسؤال الثاني الذي ينص على:" ما أثر نسب الفقـد للبيانـات (5%،10%،15%،20%).على دقة معاملات معادلة الانحدار البسيط؟"

للإجابة على السؤال الثاني: نجد أن الجدول(1-4) كذلك تدرس أثر نسب الفقد في البيانات على دقة معاملات الانحدار وتبين الجداول ظاهريا اختلاف في معامـل الثابت حسب نسب الفقد المختلفة المستخدمة في الدراسة وحسب التحليل(t-test) تدل على المعنوية الجزئية لهذه المعاملات وكذلك تحليل التباين للمعنوية الكلية للنمـوذج تدل على معنوية النموذج وأن احد قيم معاملات الانحدار تختلف عن الصفر.

للإجابة على السؤال الثالث الذي ينص على:" هل توجد فروق بين المتوسطات الحـسابية لمربعـات البـواقي لتقـدير درجـات الأفـراد تعـزى لنـسب الفقد(5%،10%،15%،20%).

تم عمل مقارنات بين المتوسطات الحسابية لمربعات البواقي للقم المولدة بدون فقد أو معالجة مع متوسطات مربعات البواقي للقيم التي فقدت وتم معالجتها بنـسب (5%،10،15%،20%). والجدول (7) يبين ذلك.

جدول رقم (7)

جدول مقارنة متوسطات مربعات البواقي بين القيم الاصلية والمعالجة بالحذف

طرق المعالجة	نسبة الفقد	الوسط الحسابي	الانحراف المعياري	قيمةT	المعنوية
	بدون فقد	3.0570	4.86790		
الحذف	5%	4.2062	9.37103	-2.112	.036
	10%	6.9212	18.70596	-3.085	.002
	15%	13.6543	26.69611	-5.629	.000
	20%	21.6002	38.97971	-2.906	.004

يتضح من الجدول أن الوسط الحسابي للبيانات يزداد بزيادة نسبة الفقد حيث أن

قيمة الوسط الحسابي للبيانات الاصلية بدون فقد (3.0570)، ثم ازدادت قيمة الوسط الحسابي عند نسبة فقد(20%) إلى (21.6002). وكذلك نلاحظ زيادة في قيمة الانحراف المعياري من (4.86790) للبيانات الاصلية ليبلغ عند نسبة فقد (20%) إلى (38.97971). ومن ذلك نجد أن طريقة المعالجة بالحذف ليست من الطرق المفضلة لمعالجة الفقد في القيم بالنسب المختلفة لأن قيمة الاحتمالية فيها أقل من (5%) وبالتالي هناك فروق بين متوسطات مربعات البواقي، وإن استخدام طريقة الحذف كطريقة معالجة للقيم المفقودة ينتج منها نتائج غير صحيحه ولا تفسر وكل ما يبنى عليها من قرارات تكون غير صحيحه.

للإجابة على السؤال الرابع الذي ينص على "هل توجد فروق بين المتوسطات الحسابية لمربعات البواقي لتقدير درجات الأفراد تعزى لطرق معالجة البيانات المفقودة(الحذف، حساب المتوسط الحسابي، التعويض المتعدد)."

جدول (8)

جدول مقارنة متوسطات مربعات البواقي بين القيم الأصلية والمعالجة بالوسط الحسابي والتعويض المتعدد

طريقة المعالجة	نسبة الفقد	الوسط الحسابي	الانحراف المعياري	قيمة t	المعنوية
	بدون فقد	3.0570	4.86790		
المتوسط الحسابي	%5	3.0931	4.88904	-1.692	.092
	%10	3.1453	5.35946	-.931	.353
	%15	3.5469	5.79637	-2.824	.005
	%20	3.8363	6.69507	-2.215	.028
التعويض المتعدد	%5	3.0724	4.85044	-.329	.742
	%10	3.0911	4.87804	-.325	.745
	%15	3.7047	6.37344	-1.900	.052
	%20	3.9730	6.13924	-1.917	.057

يتضح من الجدول أن قيمة الوسط الحسابي للقيم يزداد بنسب قليلة في كل مـن

حالة المعالجة بالمتوسط الحسابي والتعويض المتعدد، وكذلك الانحراف المعيـاري.
وطريقة المعالجة بالوسط الحساب في معالجة نسبة الفقد(5%) كانت قيمة الاحتمالية (.092) ونجد أن هذه القيمة اكبر من مستوى الدلالة (5%)وبالتالي تكون النتيجة أنه لا يوجد فروق بين متوسطات مربعات البواقي للبيانات التي تم معالجتها بالمتوسـط الحسابي عند نسبة (5%) عن بواقي القيم الاصلية؛ أما طريقة المعالجـة بالوسـط الحساب في معالجة نسبة الفقد(10%) كانت قيمة الاحتمالية (.353) ونجد أن هـذه القيمة أكبر من مستوى الدلالة(5%) وبالتالي تكون النتيجة أنه لا يوجد فروق بـين متوسطات مربعات البواقي للبيانات التي تم معالجتها بالمتوسط الحسابي عند نـسبة (10%) عن بواقي القيم الاصلية؛ اما طريقة المعالجة بالوسط الحساب في معالجة نسبة الفقد(15%) كانت قيمة الاحتمالية(.005) ونجد ان هذه القيمة أقل من مستوى الدلالة (5%) وبالتالي تكون النتيجة أنه يوجد فروق بين متوسطات مربعات البواقي للبيانات التي تم معالجتها بالمتوسط الحسابي عند نسبة (15%) عن بـواقي القـيم الاصلية وبالتالي إن المعالجة بالوسط الحسابي تكون نتائجها عند نسبة فقـد (15%) غير صحيحه على دقة معادلة الانحدار البسيط؛ أمـا طريقـة المعالجـة بالوسـط الحساب في معالجة نسبة الفقد(20%) كانت قيمة الاحتمالية(.028) ونجـد ان هـذه القيمة أقل من مستوى الدلالة (5%) وبالتالي تكون النتيجة أنه يوجـد فـروق بـين متوسطات مربعات البواقي للبيانات التي تم معالجتها بالمتوسط الحسابي عند نـسبة (20%) عن بواقي القيم الاصلية وبالتالي ان المعالجة بالوسط الحسابي تكون نتائجها عند نسبة فقد (20%) غير صحيحه على دقة معادلة الانحدار البسيط. أما طريقـة المعالجة بـالتعويض المتعـدد فـي معالجـة نـسبة الفقـد(5%) كانـت قيمـة الاحتمالية(.742) ونجد أن هذه القيمة أكبر من مستوى الدلالة (5%) وبالتالي تكون النتيجة أنه لا يوجد فروق بين متوسطات مربعات البواقي للبيانات التي تم معالجتها بالتعويض المتعدد عند نسبة (5%) عن بواقي القيم الاصلية وبالتـالي ان المعالجـة بالتعويض المتعدد تكون نتائجها عند نسبة فقد 5% صـحيحه علـى دقـة معادلـة الانحدار البسيط؛ أما طريقة المعالجة بالتعويض المتعدد في معالجـة نـسبة الفقـد (10%) كانت قيمة الاحتمالية(.745) ونجد أن هـذه القيمـة أكبـر مـن مـستوى

الدلالة(5%) وبالتالي تكون النتيجة أنه لا يوجد فروق بـين متوسـطات مربعـات البواقي للبيانات التي تم معالجتها بالتعويض المتعدد عند نسبة (10%) عن بـواقي القيم الاصلية وبالتالي إن المعالجة بالتعويض المتعدد تكون نتائجها عند نـسبة فقـد (10%) صحيحه على دقة معادلة الانحدار البسيط؛ أما طريقة المعالجة بـالتعويض المتعدد في معالجة نسبة الفقد(15%) كانت قيمة الاحتمالية (.052) ونجـد أن هـذه القيمة أكبر من مستوى الدلالة (5%) وبالتالي تكون النتيجة أنه لا يوجد فروق بـين متوسطات مربعات البواقي للبيانات التي تم معالجتها بالتعويض المتعدد عند نـسبة (15%) عن بواقي القيم الاصلية وبالتالي أن المعالجة بـالتعويض المتعـدد تكـون نتائجها عند نسبة فقد (15%) صحيحه على دقة معادلة الانحدار البسيط؛ أما طريقة المعالجة بـالتعويض المتعـدد فـي معالجـة نـسبة الفقـد(20%) كانـت قيمـة الاحتمالية(.057) ونجد أن هذه القيمة أكبر من مستوى الدلالة 5%وبالتـالي تكـون النتيجة أنه لا يوجد فروق بين متوسطات مربعات البواقي للبيانات التي تم معالجتها بالتعويض المتعدد عند نسبة (20%) عن بواقي القيم الاصلية وبالتالي إن المعالجـة بالتعويض المتعدد تكون نتائجها عند نسبة فقد (20%) صحيحه على دقـة معادلـة الانحدار البسيط. ويستنتج الباحث مما سبق أن طريقة التعويض المتعدد هي الأفضل مقارنة مع الطرق الاخرى لعدم وجود فروق ذات دلالة احصائية بـين متوسـطات مربعات البواقي للبيانات التي تم معالجتها بالتعويض المتعدد عنـد جميـع النـسب (5%،10%،15%،20%) مما يدال على عدم تأثير هذه النسب من الفقد على دقـة تقدير قيم معالم معادلة الانحدار البسيط.

2.4 التوصيات

1- إن استخدام طريقة الحذف تعد طريقة غير مفيدة ولا ينصح باستخدامها لمعالجة البيانات المفقودة كما اظهرت نتائج الدراسة أنه يوجد فـروق فـي متوسـطات مربعات البـواقي بـين البيانـات الاصـلية والبيانـات التـي فقـدت بنـسب (5%،10%،15%20%)

2- إن استخدام طريقة المعالجة بالوسط الحسابي تكون ذات نفع عند نسبة فقد (5%

و10%) والعكس عند زيادة نسبة الفقد إلى (15% و 20%)

3- ينصح باستخدام طريقة المعالجة بالتعويض المتعـدد لأفـضلية هـذه الطريقـة ونتائجها ادق في حالة وجود نسب فقد تتراوح بين(5% إلى 20%).

4- إجراء دراسة على معادلة الانحدار البسيط بنسب مختلفة وطرق معالجه مختلفة.

المراجع

أ- المراجع العربية:

اسماعيل، محمد عبدالرحمن(2001). **تحليل الانحدار الخطي البسيط**، معهد الادارة العامة، الرياض

بني عواد، ليلى (2010). **مقارنة طرق التعامل مع البيانات المفقودة في تقدير الفقرات وقدرات الافراد.** اطروحة دكتوراه غير منشورة، جامعة اليرموك الاردن.

الحيالي، علي (2013). استخدام معايير الدقة التنبؤية في تحديد الطريقة المثلى في تقدير القيم المفقودة، جامعة بغداد، كلية الزراعة قسم الاقتصاد الزراعي، **مجلة العلوم الزراعية العراقية**، ط4، ع44، ص509-517.

الدرابسة، رياض (2012). **اثر طريقة تقدير القدرة وطريقة التعامل مع القيم المفقودة على دقة تقدير معالم الفقرات والافراد.** اطروحة دكتوراه غير منشورة، جامعة اليرموك الاردن

ربيع، اسامة (2008). **التحليل الإحصائي للمتغيرات المتعددة باستخدام برنامج SPSS**.القاهرة، جامعة المنوفية

الزعبي، عمر (2013). **اثر نسبة البيانات المفقودة وطرق التعويض عنها في دقة تقدير معالم الفقرات والافراد**، اطروحة دكتوراه منشورة جامعة اليرموك الاردن

علام، صلاح الدين (2005). **نماذج الاستجابات للمفردات الاختبارية احادية البعد ومتعددة الابعاد وتطبيقاتها في القياس النفسي والتربوي**. القاهرة: دار الفكر العربي.

هاوت دينس، كرايمر دانكان(2016). **مقدمة لحزمة البرامج الإحصائية SPSS في علم النفس**، ترجمة الدكتور صلاح الدين محمود علام، دار الفكر، عمان،2015.

ب- المراجع الأجنبية:

Allison, P.D.(2006).**Imputation of categorical variables with PROC MI**. Paper presented at the annual meeting of the SAS Users Group International ,San Francisco ,CA.

Berk,Richard A.,(2003),**Regression analysis**: A constructive critique. Sage publications Inc.,p144.

Cokluk, O.& Kayri, M.(2011).**The Effects of Methods of Imputation for Missing Values on the Validity and Reliability of Scales**.Educational Sciences; Theory &Practice,11(1),303-309.

Enders, C.K.(2010).**Applied Missing Data Analysis**. New York: A Division of Guilford Publications, Inc.

Figueredo,A.J.,Mcknight,P.E.,McKnight,K.M.,&Sidani,S.(2000).**Multivariate modeling of missing data within and across assessment waves 2000,95(3),361-380.**

Gemici,S., Bendarz, A & Lim ,P .(2012). **Aprimer for handling missing Value in the analysis of education and training data,**international of Training Research,10 (3),233-250.

Graham,J.W.(2009)**.Missing data analysis : Making it work in the real world .**Annual Review of Psychology , 60, 549-576.

Hawthorne, G.& Elliott, P.(2005).**Imputing cross-sectional missing data: Comparison of common techniques.** Australian and New Zealand Journal of Psychiarty,39 (7),583-590.

Huisman,M.,Krol,B. and Van Sonderen,F.L .P.(1998). **Handling Missing Data by Re-approaching Nonrespondent Quality & Quantity.In Huisman, M. (edited) .**Item Nonresponse ; Occurrence causes ,and Imputation of Missing Answers to test Item. DSWO Press, Lieden University ,The Nethrlands,1999.

Little,R.J.A.(1988).**Robust Estimation of the Mean and Covariance Matrix from Data with Missing Values**. Applied Statistics,37,23- 380

Little,R.J.A.,& Rubin, D.B.(1987).**Statistical Analysis with Missing Data.** New York: Wiley.

Little, R. J .A.,& Rubin, D.B.(2002).**Statistical Analysis with Missing Data .** 2^{nd} **edition**. New York: John Wiley &Sons.

Ludlow .L. H.,& O'Leary ,M.(1999).**Scoring Omitted and Not-Reached Items: Practical Data Analysis Implications.** Educational and Psychological Measurment,59(4),615-630.

Lord,F,M(1980)**Application of item response theory to practical testing problems,** New Jersey: Lawrance Erlbaum Associates, Publishers.

McKnight, P.E., McKnight .K .M., Sidani ,S& Figyeredo ,A.J(2007). **Missing data: A gentle introduction.** New York: Guilford press.

Mislevy,R.J.,& Wu,P.K. (1988).**Inferring Examinee Ability When Some Item Responses are Missing (ERIC Document Reproduction Service No.ED395 017).**Priceton ,NJ: Educational Testing Service.

Peugh,J.L.,& Enders,C.K.(2004).**Missing Data in Educational Research:A Review of Reporting Practices and Suggestions for Improvement.** Review of Educational research,74(4),525 - 556.

Palta,Mari,(2003), **Quantitative methods in population health**: Extensions of ordinary regression. Wiley-IEEE,p6.

Schafer,J.L.,& Graham,J.W.(2002).**Missing data: Our view of the state of the art.** Psychological Methods,7,147-177.

Wayman,J.C.(2003).**Multiple Imputations For Missing Data:What is it and how I use it**.Paper presented at the 2003 Annual Meeting of the American Educational Research Association Chicago, IL,April.

Witta,E.L.(2002). **Effectiveness of four Methods of Handling Missing Data Using Samples From a National Database.** (ERIC Document Reproduction Service ED442810).

Witta,L.& Kaiser,J.(1991).**Four Methods of Handling Missing Data with the General Social Survey ,**Paper presented at the annual meeting of the Mid-South Educational Research Association, Lexington, KY.

معلومات شخصية

الاسم: عمران إسماعيل اللصاصمة

الكلية: العلوم التربوية/ علم النفس

التخصص: القياس والتقويم

سنة التخرج:2016

هاتف نقال:0777773303